ロバスト推定におけるバイアス-ロバストネス理論とその応用(漸近的統計理論)

安藤, 雅和; 木村, 美善

数理解析研究所講究録 1308 巻 2003 年 132-144

ロバスト推定におけるバイアス-ロバストネス理論とその応用

南山大学・経営学研究科　安藤雅和 (Masakazu Ando)
Graduate School of Business Administration, Nanzan University
南山大学・数理情報学部　木村美善 (Miyoshi Kimura)
Department of Mathematical Sciences, Nanzan University

1　はじめに

ロバスト推定は，標本の確率分布が仮定された分布（モデル分布）からずれていたり，少し異なっている場合に，モデル分布のもとで「よい」推定量がどのような影響を受けるのかを調べたり，このような場合にふさわしい推定量はどのようなものかを研究する分野である．ロバスト推定では，実際の確率分布は仮定したモデル分布とは必ずしも一致するものではなく，せいぜい近似的に等しいといえるだけである，という立場をとっている．したがって，ロバスト推定の観点からすれば，モデル分布のもとでかなりのよさを保持しつつ，仮定からのずれがあってもその影響をさほど受けず，「よさ」の損失の少ない，いわゆるロバスト（頑健な）推定量が望ましい，ということになる．ロバスト推定理論では，通常，標本の分布のモデル分布からのずれ具合を表現するためにモデル分布の近傍を用いている．近傍の種類と大きさを変えることによって，ずれ具合を表現することができる．

推定量のロバストネス（頑健性）をはかる測度としては，様々なものが提案されているが，もっとも重要なもののひとつは，推定量の近傍上での最大（あるいは最小）バイアスである．これはモデル分布の近傍上での推定量と真の母数の差の大きさの最大値を表すものであるが，推定量の大域的なロバストネスを考察するうえで有用な情報を多く持っており，ずれに対する対処応可能限界を表す破綻点（breakdown point）やモデル分布からの微小のずれに対する影響の大きさを表す GES（gross-error sensitivity）に関する情報も含んでいる．最大（あるいは最小）バイアスをロバストネスの主要な測度とし，これを最小（あるいは最大）にする推定量を求めるというアプローチ（bias robustness approach）は，Huber（1964）により提唱されたが，その重要性が近年再認識されつつあり，このアプローチからの研究が続々と発表されてきている．

位置母数の推定問題では，モデル分布が単峰で原点対称な分布の場合に，すべての位置共変推定量 (location equivariant estimator) のクラスにおいて最大バイアスを最小にする推定量はメディアンであることが Huber (1964,1981) により示された．一方，尺度（scale）母数の推定問題は，位置母数や回帰母数の推定問題における補助的な推定問題として取り扱われることが多く，主たる位置推定量あるいは回帰推定量のよさを損なわないようなものがロバスト推定量として望まれる．尺度のロバスト推定量として式が簡単で計算に時間がかからず 50% の破綻点と強いロバストネスをもつ MAD（the median absolute deviation）がよく用いられているが, Rousseeuw and Croux (1993) は MAD が

対称分布に強く依存し, 非対称分布に対しては不十分なことと漸近効率が高くない (37%) ことから，新しい 2 つの推定量 S_n, Q_n を提案し，これらが 50% 破綻点と MAD より高い漸近効率をもつことを示した.

一般の推定量については， He and Simpson (1993) により推定量の最大漸近バイアスに対する下界が与えられている．さらに，回帰推定問題では， Martin, Yohai and Zamar (1989) がロバスト線形回帰における S- 推定量と GM- 推定量の最大バイアスを導出して以来，最大バイアスを用いた回帰推定量の研究は多くの成果をあげてきている．しかし，これらの推定問題でモデル分布からのずれを表現するものとして用いられている近傍は，いずれも ε-contamination 近傍である.

本論文では，ロバスト推定におけるバイアス - ロバストネス アプローチにより，位置母数, 尺度母数 および 回帰母数の代表的なロバスト推定量の最大バイアスについて， ε-contamination 近傍を一般化した (c,γ)- 近傍上で考察する.

2 近傍

$\mathcal{X}$ を Polish 空間 (完備可分距離空間)， $\mathcal{B}$ を $\mathcal{X}$ の部分集合からなる σ- 集合体とし, $\mathcal{M}$ を $(\mathcal{X},\mathcal{B})$ 上の確率測度の全体からなる集合とする．このとき $H_0 \in \mathcal{M}$ の近傍として

$$\mathcal{P}_{H_0}(c,\gamma) = \{H \in \mathcal{M} \mid cH_0(B) - (c+\gamma-1) \le H(B) \le cH_0(B) + \gamma, {}^{\forall}B \in \mathcal{B}\}, \quad (1)$$

を考える．ここで $0 \le \gamma < 1$, $1-\gamma \le c < \infty$ とする． $c = 1-\varepsilon$, $\gamma = \varepsilon$ のときの近傍 $\mathcal{P}_{H_0}(1-\varepsilon,\varepsilon)$ は ε-contamination 近傍であり， $c = 1$, $\gamma = \delta$ のときの近傍 $\mathcal{P}_{H_0}(1,\delta)$ は total variation 近傍である．ただし， $\varepsilon, \delta \ge 0$, $\varepsilon+\delta < 1$ とする．近傍 (1) は ε-contamination と total variation を結合した近傍 (Rieder, 1977 参照) をさらに一般化したものであり，これを (c,γ)- 近傍と呼ぶことにする.

定理 2.1 $0 \le {}^{\forall}\gamma < 1$ と $1-\gamma \le {}^{\forall}c < \infty$ に対して,

$$\mathcal{P}_{H_0}(c,\gamma) = \{H = c(H_0 - W) + \gamma K \mid W \in \mathcal{W}_{H_0,\lambda}, K \in \mathcal{M}\}. \quad (2)$$

ここで $\mathcal{W}_{H_0,\lambda}$ は ${}^{\forall}B \in \mathcal{B}$ に対して $W(B) \le H_0(B)$ であり， $W(\mathcal{X}) = \lambda = (c+\gamma-1)/c$ となるようなすべての測度 W からなる集合である (Ando and Kimura, 2001a, 参照).

3 基本的結果

実数直線 R 上の確率測度 H_0 はルベーグ測度に関して単峰で原点対称な確率密度関数 f_0 をもつとし， a を H_0 の上側 $\frac{100(c+\gamma-1)}{2c}$% 点とする．また $\hat{W}$ を

$$\hat{W}(B) = H_0(B \cap [-a,\ a]^c), {}^{\forall}B \in \mathcal{B}$$

により定義し， Δ_0 は原点 0 において確率 1 をとる 1 点分布を表すとする．このとき (2) の近傍 $\mathcal{P}_{H_0}(c,\gamma)$ のもとで，確率変数の分布に関して次の定理が成り立つ.

定理 3.1 X,Y は独立で同一な分布 $H \in \mathcal{P}_{H_0}(c,\gamma)$ に従う確率変数とする．このとき，$H \times H$ のもとでの $|X-Y|$ の分布は $H = \hat{H} = c(H_0 - \hat{W}) + \gamma\Delta_0$ のとき確率的に最小になる．すなわち $0 \leq {}^{\forall}t < \infty$ に対して

$$\sup_{H \in \mathcal{P}_{H_0}(c,\gamma)} P(|X-Y| \leq t) = P_{\hat{H}\times\hat{H}}(|X-Y| \leq t) \tag{3}$$

が成り立つ．

この定理 3.1 を証明するために，次の定理を必要とする．

f を $0 < \int_{-\infty}^{\infty} f(x)dx = M < \infty$ を満たす非負の実数値関数とする．ここで M は定数である．ある正の値 $m(0 < m < M)$ に対して

$$\int_{-a}^{a} f(x)dx \;=\; m,$$

を満たす a が存在し，$\hat{g}$ を

$$\hat{g}(x) \;=\; \begin{cases} f(x), & -a \leq x \leq a, \\ 0, & \text{その他}, \end{cases}$$

とおく．さらに，

$$\begin{aligned} \mathcal{F}_0 &= \{g \in \mathcal{F} \mid 0 \leq g \leq f,\ 0 \leq \int_{-\infty}^{\infty} g(x)dx \leq m\}, \\ G(x) &= \int_{-\infty}^{x} g(t)dt, \qquad \hat{G}(x) \;=\; \int_{-\infty}^{x} \hat{g}(t)dt, \end{aligned}$$

とする．ここで $\mathcal{F}$ を R 上で定義されたすべての可測関数の集合とする．そのとき，次の定理が成り立つ．

定理 3.2 f は単峰で原点対称とする．このとき,${}^{\forall}t \geq 0$ に対して

(i) $$\sup_{g \in \mathcal{F}_0} \int_{-\infty}^{\infty} \{G(x+t) - G(x)\}g(x)dx = \int_{-\infty}^{\infty} \{\hat{G}(x+t) - \hat{G}(x)\}\hat{g}(x)dx,$$

(ii) $$\sup_{g \in \mathcal{F}_0} \int_{-\infty}^{\infty} G(x+t)g(x)dx = \int_{-\infty}^{\infty} \hat{G}(x+t)\hat{g}(x)dx,$$

が成り立つ．

4 位置母数の推定問題

H_0 を R 上の原点対称な分布で $H_\theta(x) = H_0(x-\theta)$ とし，$X_1,\ldots,X_n$ を独立で同一な分布 H に従う標本とする．このとき，$H_\theta \in \mathcal{M}$ の近傍として

$$\mathcal{P}_{H_\theta}(c,\ \gamma) = \{H \mid H(x) = c\,(H_0 - W)(x-\theta) + \gamma K(x),\ \ W \in \mathcal{W}_{H_0,\lambda}, K \in \mathcal{M}\},$$

を考える．ここでは位置母数 θ の推定量として位置共変推定量を考えるので，一般性を失うことなく $\theta=0$ と仮定することができる．このとき最大バイアスは

$$B_T(c,\gamma) \ = \ \sup\{|T(H)-T(H_0)| : H \in \mathcal{P}_{H_0}(c,\gamma)\}.$$

により定義できる．推定量のクラス $\mathcal{T}$ に対して $T^* \in \mathcal{T}$ が

$$\inf_{T\in\mathcal{T}} B_T(c,\gamma) = B_{T^*}(c,\gamma)$$

を満たすとき，T^* を $\mathcal{T}$ においてミニマックスバイアスであるという．

定理 4.1 H_0 を単峰で原点対称な密度関数 f_0 をもつ分布とし，メディアンを T_M とする．このとき，T_M はすべての移動共変推定量のクラス $\mathcal{T}$ においてミニマックスバイアスである．すなわち，

$$\inf\{B_T(c,\gamma) : T \in \mathcal{T}\} \ = \ B_{T_M}(c,\gamma).$$

この定理 4.1 は T_M がミニマックスバイアスであるという Huber の結果を拡張するものである．

5 尺度母数の推定問題

H_0 を対称で単峰な密度 f_0 をもつ分布とし，$X_1,\ldots,X_n$ を独立で同一な分布 H に従う標本とする．いま $H_{\mu,s} \in \mathcal{M}$ の近傍として

$$\begin{aligned}\mathcal{P}_{H_{\mu,s}}(c,\gamma) \ = \ & \left\{H \mid H(x) = c\,(H_0-W)\left(\frac{x-\mu}{s}\right) + \gamma K(x),\right. \\ & \left. \qquad x\in R, W \in \mathcal{W}_{H_0,\lambda}, K \in \mathcal{M}\right\}, \qquad (4)\end{aligned}$$

を考える．ここで μ は未知の位置母数，$s>0$ は未知の尺度母数であり，$\mathcal{W}_{F_0,\lambda}$, $\mathcal{M}$ は (2) で定義されたものである．s の推定量として次の 3 つを考える．

$$\begin{aligned} MAD_n &= a \operatorname*{med}_i\{|X_i - \operatorname*{med}_j X_j|\}, \\ S_n &= b \operatorname*{med}_i\{\operatorname*{med}_j |X_i - X_j|\}, \\ Q_n &= d\{|X_i - X_j| : i<j\}_{(k)}, \end{aligned}$$

ここで a, b, d はある定数である．T を尺度汎関数とするとき，近傍 (4) に関する T の最大漸近外破バイアスと最小漸近内破バイアスは，それぞれ

$$\begin{aligned} B_T^+(c,\ \gamma) &= \sup\{T(H) \ : \ H \in \mathcal{P}_{H_0}(c,\ \gamma)\}, \\ B_T^-(c,\ \gamma) &= \inf\{T(H) \ : \ H \in \mathcal{P}_{H_0}(c,\ \gamma)\}, \qquad (5) \end{aligned}$$

によって定義される．

5.1 尺度推定量の最小漸近バイアスの導出

$\mathbf{MAD}_n$ の最小漸近バイアスの導出

本節以降 X, Y は独立な確率変数を表すものとする. MAD_n の漸近形は

$$MAD(H) = a \cdot \operatorname*{med}_H\{|X - \operatorname*{med}_H Y|\}.$$

によって与えられる. このとき, MAD の内破バイアスに関して次の結果を得る.

定理 5.1 H_0 を原点対称で単峰な密度関数 f_0 をもつ分布とする. 近傍 (4) に関する MAD の内破バイアスは

$$B^-_{MAD}(c,\ \gamma) \ = \ \begin{cases} a\, H_0^{-1}\left(\dfrac{2c-2\gamma+1}{4c}\right), & if \ \ 0 \le \gamma < \frac{1}{2}, \\ 0, & if \ \ \gamma \ge \frac{1}{2}, \end{cases}$$

によって与えられる.

$\mathbf{S}_n$ の最小漸近バイアスの導出

S_n の漸近形は

$$S(H) \ = \ b \cdot \operatorname*{med}_H g_H(X),$$

によって与えられる. ここで $g_H(x)$ は

$$g_H(x) \ = \ \operatorname*{med}_H |x - Y|,$$

である. H_n が経験分布であるとき $S(H_n) = S_n$ が推定量となる. 次の定理は S の内破バイアスを与える.

定理 5.2 H_0 を原点対称で単峰な密度関数 f_0 をもつ分布とする. このとき, S の内破バイアスは

$$B^-_S(c,\ \gamma) \ = \ \begin{cases} b\, g^-\left(H_0^{-1}\left(\frac{2c-2\gamma+1}{4c}\right)\right), & \text{if} \ \ 0 \le \gamma < \frac{1}{2}, \\ 0, & \text{if} \ \ \gamma \ge \frac{1}{2}. \end{cases} \tag{6}$$

によって与えられる. ここで g^- は次式を満たす関数である.

$$H_0(x + g^-(x)) - H_0(x - g^-(x)) \ = \ \frac{1-2\gamma}{2c}. \tag{7}$$

$\mathbf{Q}_n$ の最小漸近バイアスの導出

Q_n の漸近形は

$$Q(H) = d\,G_H^{-1}\left(\frac{1}{4}\right) = d\,K_H^{-1}\left(\frac{5}{8}\right),$$

により与えられる．ここで G_H と K_H はそれぞれ H のもとでの $|X-Y|$ と $X-Y$ の分布である．K_H は原点で対称であることに注意する．次の定理は Q の内破バイアスを与える．

定理 5.3 F_0 を原点対称で単峰な密度関数 f_0 をもつ分布とする．このとき，

$$B_Q^-(c,\ \gamma) \;=\; \begin{cases} Q(\hat{H}), & \text{if } 0 \le \gamma < \frac{1}{2}, \\ 0, & \text{if } \gamma \ge \frac{1}{2}, \end{cases}$$

であり，$Q(\hat{H})$ は

$$c^2(H_0-\hat{W})^{*2}(d^{-1}\,Q(\hat{H})) + 2c\gamma(H_0-\hat{W})(d^{-1}\,Q(\hat{H})) + \gamma^2 = \frac{5}{8}, \tag{8}$$

を満たす．ここで，$\hat{H} = c\,(H_0-\hat{W}) + \gamma\Delta_0$ であり，$(H_0-\hat{W})^{*2}$ は たたみこみ (convolution)，$(H_0-\hat{W}) * (H_0-\hat{W})$ を表す．

6　ロバスト推定量の最大バイアスの下界

He and Simpson (1993) は確率分布族 $\{F_\theta\}$ における母数 θ の推定量のロバストネスを考察し，F_θ からの「ずれ」を表す ε-contamination 近傍上での推定量の最大漸近バイアスに対する下界を与えた． この下界は様々な分布族に対して有用であるが，特に θ が位置母数の場合には，メディアンがすべての位置共変推定量の中で最大バイアスを最小にするものであることを示すものともなっている．本節では，(c,γ)- 近傍上での推定量の最大漸近バイアスに対する He-Simpson タイプの下界を導出する．(c,γ)- の近傍は特殊な場合として ε-contamination 近傍を含んでおり，この結果は He and Simpson (1993) の拡張となっている．

$^\forall F,\ ^\forall G \in \mathcal{M}$ に対して，F と G の discrepancy を次のように定義する．

$$d_\varphi(G,F) = \inf\{\varphi(c,\gamma)\ :\ (c,\gamma)\in\Omega_{G,F}\}.$$

ここで,$\varphi(c,\ \gamma)$ は非減少，連続で，$\varphi(1,0)=0$ を満たす非負実数値関数であり，

$$\begin{aligned} \Omega_{G,F} &= \{(c,\gamma)\in\Omega\ :\ G(B) \le c\,F(B)+\gamma,\ \ ^\forall B\in\mathcal{B}\}, \\ \Omega &= \{(c,\gamma)\ :\ 1-\gamma \le c < \infty,\ 0 \le \gamma < 1/2\}, \end{aligned}$$

とする．この d_φ を用いて discrepancy a の F の近傍を

$$\mathcal{P}_F^\varphi(a) = \{G \in \mathcal{M} \mid d_\varphi(G, F) \le a\},$$

により定義する．近傍は次のようになることがわかる．

$$\mathcal{P}_F^\varphi(a) \;=\; \bigcup_{\varphi(c,\gamma)\le a} \mathcal{P}_F(c,\ \gamma).$$

このとき，$\mathcal{P}_{F_\theta}^\varphi(a)$ 上での θ の推定汎関数 ($\mathcal{M}$ から パラメータ空間 Θ への写像) T の最大バイアスは

$$b_T^\varphi(a,\ F_\theta) = \sup\{\rho(T(G),\ \theta)\ :\ G \in \mathcal{P}_{F_\theta}^\varphi(a)\}$$

となる．ここで ρ は Θ 上の距離を表す．

6.1 一般パラメータ族の場合

任意の (c,γ) と任意の $W \in \mathcal{W}_{F_0,\lambda}$ に対して improper 分布のパラメータ族 $\{\tilde{F}_{W,\theta}\}$, $\tilde{F}_{W,\theta} = (F_0 - W)_\theta$, を考える．

$\tilde{F}_{\theta,W}$ と $\tilde{F}_{\eta,W}$ の variation distance は

$$\tilde{d}_v(\tilde{F}_{\theta,W}, \tilde{F}_{\eta,W}) \;=\; \sup\{|\tilde{F}_{\theta,W}(B) - \tilde{F}_{\eta,W}(B)|\ :\ B \in \mathcal{B}\}. \tag{9}$$

により与えられる．このとき， Donoho and Liu (1988) のように variation gauge $\tilde{b}_{v,W}$ を

$$\tilde{b}_{v,W}(a,\ F_\theta) \;=\; \sup\{\rho(\theta,\ \eta)\ :\ \tilde{d}_v(\tilde{F}_{\theta,W}, \tilde{F}_{\eta,W}) \le a \text{ を満たす } \eta\}. \tag{10}$$

によって定義する．

定理 6.1 $\{F_\theta\}$ を σ-有限測度 μ に関して絶対連続であるとし，(c_0, γ_0) を Ω の点とする．T が θ の推定汎関数ならば，任意の $W \in \mathcal{W}_{F_0,\lambda}$ に対して

$$\sup_{\eta:\rho(\theta,\eta)\le\tilde{b}_{v,W}((1-\lambda)\frac{a}{1+a},F_\theta)} b_T^\varphi(J_\lambda(a),\ F_\eta) \;\ge\; \frac{1}{2}\tilde{b}_{v,W}\left((1-\lambda)\frac{a}{1+a},\ F_\theta\right),\quad a \ge 0, \tag{11}$$

が成り立つ．ここで，

$$\begin{aligned} J_\lambda(a) &= \varphi(c^*(a),\ \gamma^*(a)),\quad \lambda = \frac{c_0+\gamma_0-1}{c_0}, \\ c^*(a) &= \frac{1+a}{(1-\lambda)(1+2a)},\quad \gamma^*(a) = \frac{a}{1+2a}. \end{aligned}$$

$\mathcal{P}_F(c,\gamma)$ 上での T の最大バイアス $B_T(c,\gamma;F)$ の下界を評価するために，$\frac{1}{2}<c\leq 1$ と $c\geq 1$ に分けて考える．はじめに $\frac{1}{2}\leq c\leq 1$ の場合を取り上げる．ここで，$\Omega_1\subset\Omega$ を

$$\Omega_1=\left\{(c,\gamma)\ :\ 1-\gamma\leq c\leq 1,\ 0\leq\gamma<\frac{1}{2}\right\}.$$

とし，$\varphi_1(c,\gamma)$ を

$$\varphi_1(c,\gamma)=\varphi_{k,\lambda}^{(1)}(c,\gamma)=\max(1-c,\ k(c+\gamma-1)), \tag{12}$$

とする．k は 正の実数である．このとき，$J_1(\xi)$ は

$$J_1(\xi)=J_{k,\lambda}^{(1)}(\xi)=\varphi_{k,\lambda}^{(1)}(c^*(\xi),\gamma^*(\xi))=\max(1-c^*(\xi),\ k(c^*(\xi)+\gamma^*(\xi)-1)),$$

となる．さらに

$$\xi\ \geq\ \frac{(k+1)\lambda}{1-(k+2)\lambda},\quad 0<k\leq\frac{1-2\lambda}{\lambda},\quad \lambda=\frac{c_1+\gamma_1-1}{c_1}. \tag{13}$$

を満たすので，

$$\mathcal{P}_F^{\varphi_1}(a)=\bigcup_{\varphi_1(c,\gamma)\leq a}\mathcal{P}_F(c,\gamma)=\mathcal{P}_F\left(1-a,\frac{(k+1)a}{k}\right),$$

から

$$b_T^{\varphi_1}(J_1(a),F_\eta)=B_T\left(1-J_1(a),\left(\frac{k+1}{k}\right)J_1(a);F_\eta\right).$$

がいえる．それゆえ次の定理が成り立つ．

定理 6.2 (c_1,γ_1) を Ω_1 の点とする．T が θ の推定汎関数ならば，任意の $W\in\mathcal{W}_{F_0,\lambda}$ に対して

$$\sup_{\eta:\rho(\theta,\eta)\leq\tilde{b}_{v,W_\lambda}((1-\lambda)\frac{a}{1+a},F_\theta)} B_T\left(1-J_1(a),\left(\frac{k+1}{k}\right)J_1(a);F_\eta\right)\geq\frac{1}{2}\tilde{b}_{v,W_\lambda}\left((1-\lambda)\frac{a}{1+a},F_\theta\right),$$

が成り立つ．

次に $c\geq 1$ の場合に最大バイアス $B_T(c,\gamma;F)$ の下界を導出する．$\Omega_2,\varphi_2(c,\gamma),J_2(\xi)$ を次のようにとる．

$$\begin{aligned}\Omega_2 &= \left\{(c,\gamma)\ :\ 1\leq c<\infty,\ 0\leq\gamma<\frac{1}{2}\right\},\\ \varphi_2(c,\gamma) &= \varphi_{k,\lambda}^{(2)}(c,\gamma)=\max(c-1,k\gamma-1),\\ J_2(\xi) &= J_{k,\lambda}^{(2)}(\xi)=\varphi_{k,\lambda}^{(2)}(c^*(\xi),\gamma^*(\xi))=\max(c^*(\xi)-1,k\gamma^*(\xi)-1).\end{aligned}$$

$$\xi \geq \frac{1}{(1-\lambda)k-1}, \quad 0 < k \leq \frac{1}{1-\lambda}, \quad \lambda = \frac{c_2+\gamma_2-1}{c_2}. \tag{14}$$

を満たすので，

$$\mathcal{P}_F^{\varphi_2}(a) = \bigcup_{\varphi_2(c,\gamma)\leq a} \mathcal{P}_F(c,\gamma) = \mathcal{P}_F\left(a+1,\ \frac{a+1}{k}\right),$$

から

$$b_T^{\varphi_2}(J_2(a),\ F_\eta) \ = \ B_T\left(J_2(a)+1,\ \frac{1}{k}(J_2(a)+1);\ F_\eta\right). \tag{15}$$

がいえる．それゆえ，次の定理が成り立つ．

定理 6.3 (c_2,γ_2) を Ω_2 の点とする．T が θ の推定汎関数ならば，任意の $W \in \mathcal{W}_{F_0,\lambda}$ に対して

$$\sup_{\eta\ :\ \rho(\theta,\eta)\leq \tilde{b}_{v,W_\lambda}\left((1-\lambda)\frac{a}{1+a},\ F_\theta\right)} B_T\left(J_2(a)+1,\ \frac{1}{k}(J_2(a)+1);\ F_\eta\right) \geq \frac{1}{2}\tilde{b}_{v,W_\lambda}\left((1-\lambda)\frac{a}{1+a},\ F_\theta\right), \tag{16}$$

が成り立つ．

7 回帰母数の推定問題

線形回帰モデル

$$y_i \ = \ \boldsymbol{\theta}_0'\boldsymbol{x}_i + \beta_0 + u_i, \quad i = 1,\cdots n$$

を考える．ここで $\boldsymbol{x}_i = (x_{i1},\ldots,x_{ip})'$ を p 次元ユークリッド空間 R^p の値をとる確率ベクトル，$\boldsymbol{\theta}_0 = (\theta_{01},\ldots,\theta_{0p})'$ を回帰母数ベクトル，β_0 を切片，u_i は独立な誤差で $\boldsymbol{x}_i$ とは独立な確率変数とする．F_0 を u_i の 分布関数，G_0 を $\boldsymbol{x}_i$ の 分布関数，H_0 を $(y_i,\boldsymbol{x}_i)$ の 分布関数とする．また，$\boldsymbol{T}$ を R^{p+1} 上の分布関数 H の全体からなる集合 $\mathcal{H}$ 上で定義された $\boldsymbol{\theta}_0$ の汎関数とする．H における $\boldsymbol{T}$ のバイアスとして

$$b(\boldsymbol{T},H) = \{[\boldsymbol{T}(H)-\boldsymbol{\theta}_0]'\Sigma_0[\boldsymbol{T}(H)-\boldsymbol{\theta}_0]\}^{1/2}$$

を考える．ただし，Σ_0 はアフィン共変分散共分散行列である．回帰共変な $\boldsymbol{T}$ のみを扱うので，$b(\boldsymbol{T},H)$ の不変性により，一般性を失うことなく $\boldsymbol{\theta}_0 = \boldsymbol{0}$, $\Sigma_0 = I$ とすることができ，$b(\boldsymbol{T},H) = \|\boldsymbol{T}(H)\|$ となる．このとき，$\mathcal{P}_{H_0}(c,\gamma)$ 上での $\boldsymbol{T}$ の最大漸近バイアスは

$$B_{\boldsymbol{T}}(c,\gamma) = \sup\{b(\boldsymbol{T},H)\ :\ H \in \mathcal{P}_{H_0}(c,\gamma)\}$$

であり， $\boldsymbol{T}$ は漸近不偏 すなわち $\boldsymbol{T}(H_0)=\boldsymbol{0}$ と仮定する．

一方、 H における 切片 T_0 のバイアスは

$$b(H)=|T_0(H)-\beta_0+(\boldsymbol{T}(H)-\boldsymbol{\theta}_0)'\boldsymbol{\mu}_0|,$$

と表すことができる．ここで $\boldsymbol{\mu}_0$ は $\boldsymbol{x}$ の G_0 のもとでの多変量位置母数である．この場合も同様に 一般性を失うことなく $\boldsymbol{\theta}_0=\boldsymbol{\mu}_0=\boldsymbol{0}$, $\beta_0=0$ と仮定できるので, バイアスは

$$b(H)=|T_0(H)|$$

となり， $\mathcal{P}_{H_0}(c,\gamma)$ 上での T_0 の最大漸近バイアスは

$$B_{T_0}(c,\gamma)=\sup_{H\in\mathcal{P}_{H_0}(c,\gamma)}|T_0(H)|$$

となる．

次のように定義される推定汎関数 $T_0,\boldsymbol{T}$ のクラスを考える:

$$[T_0(H),\boldsymbol{T}(H)]=\arg\min_{\beta,\boldsymbol{\theta}}J(F_{H,\beta,\boldsymbol{\theta}}), \tag{17}$$

ここで $J(\cdot)$ をロバスト損失汎関数， $F_{H,\beta,\boldsymbol{\theta}}$ を分布 H のもとでの残差の絶対値 $r_i(\beta,\boldsymbol{\theta})=|y_i-\beta-\boldsymbol{\theta}'\boldsymbol{x}_i|$ の分布関数とする．この汎関数のクラスには， S 推定量, τ 推定量, R 推定量など様々なロバスト推定量が含まれる (Berrendero and Zamar, 2001 参照)．分布 H からの 標本 $(y_1,\boldsymbol{x}_1),\ldots,(y_n,\boldsymbol{x}_n)$ の経験分布を H_n とするとき， $\boldsymbol{T}_n=\boldsymbol{T}(H_n)$ が $\boldsymbol{\theta}_0$ の推定量となる． $\boldsymbol{T}$ と T_0 が $\mathcal{P}_{H_0}(c,\gamma)$ 上で残差許容的 (residual admissible) であるとは, $(0,\infty)$ 上で連続な分布関数

$$F_1(v)\ <\ F_2(v),\qquad {}^{\forall}v\geq 0$$

に対して， $(0,\infty)$ 上で連続な分布関数 $F_{H_n,T_0(H_n),\boldsymbol{T}(H_n)}(v)$, $F_{H_n,\beta^*,\boldsymbol{\theta}^*}(v)$ で

$$\begin{aligned}\lim_{n\to\infty}F_{H_n,T_0(H_n),\boldsymbol{T}(H_n)}(v) &= F_1(v),\quad {}^{\forall}v>0\\ \lim_{n\to\infty}F_{H_n,\beta^*,\boldsymbol{\theta}^*}(v) &= F_2(v),\quad {}^{\forall}v>0\end{aligned}$$

を満たすような，分布列 $\{H_n\}\subset\mathcal{P}_{H_0}(c,\gamma)$, $\boldsymbol{\theta}^*\in R^p$ と $\beta^*\in R$ が存在しないことをいう (Yohai and Zamar, 1993 参照)．

$\varphi=\{W_{\beta,\boldsymbol{\theta}}\in\mathcal{W}_{H_0,\lambda}:\beta\in R,\ \boldsymbol{\theta}\in R^p\}$ を $H_0-W_{\beta,\boldsymbol{\theta}}$ のもとでの $|y-\beta-\boldsymbol{\theta}'\boldsymbol{x}|$ の分布 $F_{\varphi,\beta,\boldsymbol{\theta}}(=F_{(H_0-W_{\beta,\theta}),\beta,\boldsymbol{\theta}})$ が β と $\|\boldsymbol{\theta}\|$ に依存するような測度 $W_{\beta,\boldsymbol{\theta}}$ からなる族とし, φ の全体からなる集合を $\mathcal{F}$ とする．最大バイアスの上界の導出に必要な $\hat{\varphi}=\{\hat{W}_{\beta,\theta}\}$ と $\varphi^*=\{W^*_{\beta,\theta}\}$ を次のように定義する．

$$\begin{aligned}\hat{W}_{\beta,\theta}(B) &= H_0\left(B\cap\left\{|y-\beta-\boldsymbol{\theta}'\boldsymbol{x}|\geq a_{\beta,\|\theta\|}\left(\frac{c+\gamma-1}{c}\right)\right\}\right),\quad {}^{\forall}B\in\mathcal{B}^{p+1},\\ W^*_{\beta,\theta}(B) &= H_0\left(B\cap\left\{|y-\beta-\boldsymbol{\theta}'\boldsymbol{x}|\leq a_{\beta,\|\theta\|}\left(\frac{1-\gamma}{c}\right)\right\}\right),\quad {}^{\forall}B\in\mathcal{B}^{p+1},\end{aligned}$$

ここで $a_{\beta,\|\theta\|}(\eta)$ $(0<\eta<1)$ は $|y-\beta-\theta' x|$ の分布の上側 $100\eta\%$ 点, すなわち,

$$H_0\left(|y-\beta-\boldsymbol{\theta}'\boldsymbol{x}| \geq a_{\beta,\|\theta\|}(\eta)\right)=\eta.$$

また, $d_\varphi = J[c\,F_{\varphi,0,\mathbf{0}}+\gamma\delta_\infty]$ とし,

$$m_\varphi(t)=\inf_{\|\boldsymbol{\theta}\|=t}\inf_{\beta\in R} J[c\,F_{\varphi,\beta,\boldsymbol{\theta}}+\gamma\delta_0], \tag{18}$$

とおく. ここで, δ_0,δ_∞ は, それぞれ 0 と ∞ で確率 1 をもつ分布関数である.

ロバスト損失汎関数は次の仮定を満たすとする.

A1 (a) F,G は $[0,\infty)$ 上の分布関数で, $\forall u\geq 0$ に対して $F(u)\leq G(u)$ を満たすならば

$$J(F)\geq J(G).$$

(b) $\{F_n\}$,$\{G_n\}$ は $[0,\infty)$ 上の分布関数列で $(0,\infty)$ 上で連続とし, $F_n(u)\to F(u)$ かつ $G_n(u)\to G(u)$ とする. また, F,G は $(0,\infty)$ 上の分布関数 (全確率が 1 より小さい場合もあり得る) で連続, $G(\infty)\geq 1-\varepsilon$ で $G(u)\geq F(u),\forall u\geq 0$ とする. このとき

$$\lim_{n\to\infty} J(F_n)\geq \lim_{n\to\infty} J(G_n)$$

が成り立つ. さらに, 仮定の不等式が狭義ならば結果も狭義で成り立つ.

(c) F,G は $[0,\infty)$ 上の分布関数とし, F は連続とする. このとき

$$\begin{aligned} J(c\,F_{H-W^*}+\gamma\delta_\infty) &\geq J(c\,F_{H-W}+\gamma\delta_\infty)\approx \lim_{n\to\infty} J(c\,F_{H-W}+\gamma U_n)\\ &\geq J(c\,F_{H-W}+\gamma G)\geq J(c\,F_{H-\hat{W}}+\gamma G).\end{aligned}$$

ここで U_n は $[n-\frac{1}{n},n+\frac{1}{n}]$ 上の一様分布に従う.

A2. F_0 は原点に関して対称で単峰な連続密度関数 f_0 $(f_0(u)>0,{}^\forall u\in R)$ をもち, $P_{G_0}(\boldsymbol{x}'\boldsymbol{\theta}=c_1)<1$, ${}^\forall\boldsymbol{\theta}\in R^p(\boldsymbol{\theta}\neq 0)$, $c_1\in R$ を満たす.

補題 7.1 A1(b) と A2 のもとで

$$J(c\,F_{\varphi,\beta(\boldsymbol{\theta}),\boldsymbol{\theta}}+\gamma\delta_0)=\inf_{\beta\in R} J(c\,F_{\varphi,\beta,\boldsymbol{\theta}}+\gamma\delta_0)$$

を満たす $\beta(\boldsymbol{\theta})\in R$ が存在する. さらに, ${}^\forall t>0,{}^\forall\boldsymbol{\theta}\in\{\boldsymbol{\theta}:\|\boldsymbol{\theta}\|=t\}$ に対して $|\beta(\boldsymbol{\theta})|\leq K_t$ となる $K_t>0$ が存在する.

補題 7.2 A2 のもとで, すべての $\|\boldsymbol{\theta}\|=1,\lambda>0,u>0$ に対して, $F_{\varphi,\lambda\beta,\lambda\boldsymbol{\theta}}(u)$ は λ に関して狭義減少である.

補題 7.3 $m_\varphi(t)$ を式 (18) により定義されたものとする. このとき, A1(b) と A2 の仮定のもとで, 次が成り立つ.

(a) $\|\boldsymbol{\theta}_t\| = t,\ \ m_\varphi(t) = J(c\,F_{\varphi,\beta(\boldsymbol{\theta}_t),\boldsymbol{\theta}_t} + \gamma\delta_0)$ を満たす $\boldsymbol{\theta}_t \in R^p$ と $\beta(\boldsymbol{\theta}_t) \in R$ が存在する.

(b) $m_\varphi(t)$ は狭義増加関数である.

これらの補題により次の定理を得る.

定理 7.1 $\boldsymbol{T}$ を (17) により定義されたものとする. このとき

$$\underline{B}_{\boldsymbol{T}}(c,\gamma) \le B_{\boldsymbol{T}}(c,\gamma) \le \overline{B}_{\boldsymbol{T}}(c,\gamma), \qquad (\gamma < \min(b, 1-b)\text{のとき}),$$

$$B_{\boldsymbol{T}}(c,\gamma) = \infty, \qquad (\gamma \ge \min(b, 1-b)\text{のとき}).$$

ただし

$$\overline{B}_{\boldsymbol{T}}(c,\gamma) = m_{\tilde{\varphi}}^{-1}(d_{\varphi^*}), \quad \underline{B}_{\boldsymbol{T}}(c,\gamma) = \sup_{\varphi\in\mathcal{F}} m_\varphi^{-1}(d_\varphi).$$

参考文献

[1] Ando,M. and Kimura,M.(1999) On the maximum bias of the least α-quantile estimators for robust regression over neighborhoods defined by special capacities, *Nanzan Management Review*, **14**, 383-396. (in Janapnese)

[2] Ando,M. and Kimura,M.(2001a). A characterization of the neighborhoods defined by certain special capacities and their applications to bias-robustness of estimates, *J. Statist. Plann. Inference*. To appear.

[3] Ando,M. and Kimura,M.(2001b) The maximum asymptotic bias of S-estimates for regression over the neighborhoods defined by certain special capacities, Technical Report NANZAN-TR-2001-02, *Nanzan Academic Society*, Submitted.

[4] Berrendero,J.R. and Zamar,R.H.(2001). Maximum bias curves for robust regression with non-elliptical regressors, *Ann. Statist.*, **29**, 224-251.

[5] Chen,Z.(1998). A note on bias robustness of the median, *Statist. Probab. Lett.*, **38**, 363-368.

[6] Davies, P.L.(1990). Asymptotics of S-estimators in the linear regression models. *Ann. Statist.*, **18**, 1651-1675.

[7] He,X. and Simpson,D.G.(1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation, *Ann. Statist.*, **21**, 314-337.

[8] Hössjer, O.(1992). On the optimality of S-estimators. *Statistics and Probability Letters*, **14**, 413-419.

[9] Huber,P.J.(1964). Robust estimation of a location parameter, *Ann. Math. Statist.*, **35**, 73-101.

[10] Huber,P.J.(1981). *Robust Statistics*, Wiley, New York.

[11] Martin,R.D., Yohai,V.J. and Zamar,R.H.(1989). Min-max bias robust regression, *Ann. Statist.*, **17**, 1608-1630.

[12] Rieder,H.(1977). Least favorable pairs for special capacities. *Ann. Statist.*, **6**, 1080-1094.

[13] Rousseeuw, P.J.(1984). Least median of squares regression, *J. Amer. Statist. Assoc.* **79**, 871-880.

[14] Rousseeuw, P.J. and Yohai, V.(1984). Robust regression by means of S-estimators, *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist*, **26**, Springer, New York, 256-272.

[15] Rousseeuw, P.J. and Croux, C.(1993). Alternatives to the median absolute deviation, *J. Amer. Statist. Assoc.*, **88**, 1273-1283.

[16] Yohai,V.J. and Zamar,R.H.(1993). A minimax-bias property of the least α-quantile estimates, *Ann. Statist.*, **21**, 1824-1842.